

# Hodnocení vedoucího závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

**Student:** Bc. Markéta Filipová  
**Vedoucí práce:** Ing. Jaroslav Kuchař  
**Název práce:** Nástroj pro analýzu článků z českých zpravodajských serverů  
**Obor:** Webové a softwarové inženýrství

**Datum vytvoření:** 30. 5. 2016

<b>Hodnotící kritérium:</b> <b>1. Náročnost a další komentář k zadání</b>	<b>Způsob hodnocení - následující škálou 1 až 5:</b> <b>1=mimořádně náročné zadání,</b> <b>2=náročnější zadání,</b> <b>3=průměrně náročné zadání,</b> <b>4=lehčí, ale ještě dostatečně náročné zadání,</b> <b>5=nedostatečně náročné zadání</b>
<b>Popis kritéria:</b> Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.) <b>Komentář:</b> Cílem ZP je seznámit se problematikou získávání obsahu z webu, jeho zpracováním a následnou analýzou s využitím metod zpracování a analýzy textu. Práce vyžaduje znalost poměrně široké oblasti teorie i množství souvisejících technologií. ZP navazuje na poslední výsledky z výzkumu i praxe. Vznikla i drobná rozšíření oproti zadání jako je mapování detekovaných entit v textu do Linked Open Data s využitím české varianty DBpedia. Zadání vzhledem k širokému pokrytí považuji za obtížnější.	
<b>Hodnotící kritérium:</b> <b>2. Splnění zadání</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b> <b>1=zadání splněno,</b> <b>2=zadání splněno s menšími výhradami,</b> <b>3=zadání splněno s většími výhradami,</b> <b>4=zadání nesplněno</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků. <b>Komentář:</b> Zadání bylo splněno v plném rozsahu. V rámci práce vzniklo kompletní řešení pro získání článků, extrakci relevantního obsahu a jeho analýzu. Vše zaměřeno na prostředí českého webu a jazyka. Zejména algoritmus pro extrakci informací z HTML a následná textová analýza jsou podrobně navrženy a zpracovány.	
<b>Hodnotící kritérium:</b> <b>3. Rozsah písemné zprávy</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b> <b>1=splňuje požadavky,</b> <b>2=splňuje požadavky s menšími výhradami,</b> <b>3=splňuje požadavky s většími výhradami,</b> <b>4=nesplňuje požadavky</b>
<b>Popis kritéria:</b> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. <b>Komentář:</b> Práce splňuje požadavky na rozsah. Práce zahrnuje vše od teoretických základů, přes návrhy, implementace a testování dílčích částí po výsledný experiment na vybraných českých zpravodajských serverech. ZP neobsahuje zbytečné části.	
<b>Hodnotící kritérium:</b> <b>4. Věcná a logická úroveň práce</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b> <b>95 (A)</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. <b>Komentář:</b> ZP je po věcné stránce v pořádku. ZP je logicky členěna do kapitol odpovídajících dílčím částem zadání a vlastního řešení. Kapitoly na sebe vhodně navazují. Čitelnosti a pochopitelnosti by pomohlo více příkladů při vysvětlování navrženého algoritmu.	
<b>Hodnotící kritérium:</b> <b>5. Formální úroveň práce</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b> <b>95 (A)</b>
<b>Popis kritéria:</b> Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.	

**Komentář:**

ZP je po formální i jazykové stránce v pořádku. Drobnou výtka jsou v některých částech česko-anglické kombinace pojmů.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**6. Práce se zdroji**

95 (A)

**Popis kritéria:**

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a uvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

**Komentář:**

V práci jsou využity elektronické i odborné zdroje. Práce se zdroji je v pořádku. Použitá řešení jsou v souladu s licenčními podmínkami a autorským právem.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**7. Hodnocení výsledků, publikační výstupy a ocenění**

95 (A)

**Popis kritéria:**

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

**Komentář:**

V rámci ZP vznikla nová metoda/algoritmus pro extrakci relevantních dat z HTML. Algoritmus je zaměřen na české zpravodajské články. Je inspirován publikovanými metodami a kombinuje několik přístupů. Dále vzniklo řešení pro extrakci entit a analýzu sentimentu stavějící a rozšiřující existujících nástroje.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

**8. Komentář o využitelnosti výsledků**

**Popis kritéria:**

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

**Komentář:**

Výsledkem práce je zároveň sada samostatných modulů implementovaných v node.js. Jednotlivé moduly jsou využitelné opakovaně samostatně a nezávisle. Nejedná se tedy o jednoúčelové řešení.

Hodnotící kritérium:

Způsob hodnocení - následující škálou 1 až 5:

**9. Aktivita a samostatnost studenta v průběhu řešení**

9a:

**1=výborná aktivita,**

**2=velmi dobrá aktivita,**

**3=průměrná aktivita,**

**4=slabší, ale ještě dostatečná aktivita,**

**5=nedostatečná aktivita**

9b:

**1=výborná samostatnost,**

**2=velmi dobrá samostatnost,**

**3=průměrná samostatnost,**

**4=slabší, ale ještě dostatečná samostatnost,**

**5=nedostatečná samostatnost**

**Popis kritéria:**

Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (9a). Posuďte schopnost studenta samostatné tvůrčí práce (9b).

**Komentář:**

Studentka prokázala výbornou samostatnou a aktivní práci na zadaném problému.

Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

**10. Celkové hodnocení**

95 (A)

**Popis kritéria:**

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

**Text hodnocení:**

Studentka se seznámila s širokou teorií i množstvím technologií potřebných pro zpracování zadání ZP. ZP je zaměřena na získávání zpravodajských článků, extrakci relevantních informací (text, titulky, datum publikace a případná klíčová slova) a následnou analýzu článků. Analýza zahrnovala především seznámení se s problematikou analýzy textu s využitím detekce entit a analýzou sentimentu. Vznikla i drobná rozšíření oproti zadání jako je mapování detekovaných entit v textu do Linked Open Data s využitím české varianty DBpedia V rámci ZP vznikla sada samostatných a znovupoužitelných modulů pro Node.js. Doporučuji k obhajobě.

Podpis vedoucího práce: